Course Code: 20MC9141

R20

SIDDHARTH INSTITUTE OF ENGINEERING & TECHNOLOGY::PUTTUR
(AUTONOMOUS)
Siddharth Nagar, Narayanavanam Road – 517583

QUESTION BANK (DESCRIPTIVE)

Subject with Code : Big Data Analytics (20MC9141) **Course & Branch**: MCA

Year & Sem: II-MCA & II-Sem **Regulation:** R20

UNIT – I

INTRODUCTION TO BIG DATA & STATISTICAL CONCEPTS

1		Generalize the following in detail. Challenges of Conventional System ii) Nature of Data	**[L2][CO1] [12M]**
2	**a)**	Discuss about intelligent data analysis and nature of data.	**[L2][CO1] [06M]**
	b)	Explain types of Data in Big Data Analytics.	**[L2][CO1] [06M]**
3		Explain about analytic processes and tools in big data.	**[L2][CO1] [12M]**
4		Define and explain Statistical Inference with an example.	**[L1][CO6] [12M]**
5	**a)**	What are the different inferences in big data analytics?	**[L1][CO1] [06M]**
	b)	Derive the statistical inference for the following. Question: From the shuffled pack of cards, a card is drawn. This trial is repeated for 400 times, and the suits are given below:	**[L3][CO6] [06M]**

Suit	Spade	Clubs	Heart	Diamnonds
No of times drawn	90	100	120	90

While a card is tried at random, then what is the probability of getting a Diamond cards, Black cards, Except for spade

6	**a)**	What is bootstrapping? outline its importance.	**[L2][CO6] [06M]**
	b)	List and Explain the Characteristics of big data	**[L2][CO1] [06M]**
7	**a)**	What is sampling and sampling distribution give a detailed analysis.	**[L4][CO6] [06M]**
	b)	Clearly explain about modern data analytic tools.	**[L2][CO1] [06M]**
8		Define and explain the following. i) Intelligent Data Analysis ii) Analysis Vs Reporting.	**[L2][CO1] [12M]**
9		Demonstrate Prediction Error with an example.	**[L2][CO6] [12M]**
10	**a)**	Discuss any five characteristics of Big Data.	**[L2][CO1] [06M]**
	b)	Analyze the concept of resampling in big data.	**[L4][CO6] [06M]**

Course Code: 20MC9141

R20

UNIT –II

INTRODUCTION TO STREAM CONCEPTS & REAL TIME ANALYTICS PLATFORM

1	**a)**	What is a data stream? Explain the nature of data steam	**[L2][CO1] [06M]**
	b)	Discuss the benefits of stream data processing.	**[L2][CO1] [06M]**
2	**a)**	Explain the different applications of data streams in detail.	**[L2][CO5] [06M]**
	b)	Summarize the real time applications of stream computing.	**[L2][CO5] [06M]**
3	**a)**	Clearly, explain the stream model and architecture.	**[L2][CO1] [06M]**
	b)	Examine the sources of data steams.	**[L3][CO1] [06M]**
4		Demonstrate the counting of ones in a window using DGIM algorithm with an example.	**[L2][CO2] [12M]**
5		Discuss in detail Counting distinct elements in a stream with example.	**[L2][CO2] [12M]**
6	**a)**	What are filters in Big Data? Explain Bloom Filter with example	**[L1][CO1] [06M]**
	b)	Explain about sampling data in a stream.	**[L2][CO1] [06M]**
7	**a)**	What are the issues of stream processing?	**[L1][CO1] [06M]**
	b)	Define about stock market predictions.	**[L1][CO5] [06M]**
8		Explain the following. i) FM algorithm and its application ii) AMS algorithm and its applications	**[L2][CO2] [12M]**
9	**a)**	What is Real Time Analytics? Discuss about RTAP applications.	**[L2][CO5] [06M]**
	b)	Illustrate Real Time Sentiment Analysis in real time analytics platform.	**[L3][CO5] [06M]**
10	**a)**	Discuss in detail about estimating moments with an example.	**[L2][CO1] [06M]**
	b)	Generalize stream concepts in big data analytics.	**[L2][CO1] [06M]**

Course Code: 20MC9141 **R20**

UNIT-III

HISTORY OF HADOOP & DEVELOPING A MAP REDUCE APPLICATION

1	**a)**	What is Hadoop? Explain its components.	**[L1][CO3] [06M]**
	b)	How do you analyze the data in hadoop?	**[L1][CO3] [06M]**
2		Discuss in detail designing of HDFS.	**[L2][CO6] [12M]**
3	**a)**	List and explain different failures in Map reduce.	**[L2][CO4] [06M]**
	b)	List out Big Data Analytical Tools.	**[L1][CO5] [60M]**
4		Explain the Anatomy of MapReduce Job.	**[L2][CO4] [12M]**
5	**a)**	How Hadoop streaming is suited with text processing? Explain.	**[L1][CO4] [06M]**
	b)	Discuss java interfaces to HDFS basics.	**[L2][CO6] [06M]**
6		Explain the following i) History of Hadoop. ii) Task Execution	**[L2][CO3] [12M]**
7	**a)**	Analyze the concept of developing the Map Reduce Application.	**[L4][CO4] [06M]**
	b)	How failures are handled in Map Reduce Applications?	**[L1][CO3] [06M]**
8	**a)**	How map reduce job works with classic MapReduce?	**[L1][CO4] [06M]**
	b)	Describe job scheduling.	**[L2][CO6] [06M]**
9	**a)**	Clearly explain how map reduce jobs run on YARN.	**[L2][CO4] [06M]**
	b)	Briefly explain about shuffling and sorting.	**[L2][CO3] [06M]**
10		Discuss the various types of MapReduce& its formats.	**[L2][CO3] [12M]**

Course Code: 20MC9141 **R20**

UNIT-IV

SETTING UP HADOOP CLUSTER & ADMINISTERING HADOOP

1	**a)**	What is Cluster? Explain the setting up a Hadoop cluster.	**[L1][CO3] [06M]**
	b)	Explain about the procedure of Cluster setup and Installation.	**[L2][CO3] [06M]**
2	**a)**	What are the different types of Hadoop configuration files? Discuss.	**[L1][CO3] [06M]**
	b)	Summarize the concept of Administering Hadoop.	**[L2][CO3] [06M]**
3	**a)**	What are the additional configuration properties to set for HDFS?	**[L1][CO3] [06M]**
	b)	Discuss about Cluster specification.	**[L2][CO6] [06M]**
4		Briefly discuss about Monitoring and maintenance in Hadoop system administration.	**[L2][CO3] [12M]**
5		Explain the following i) HDFS Monitoring ii) Hadoop in the cloud	**[L2][CO1] [12M]**
6		How will you define commissioning new nodes and decommissioning oldnodes.	**[L1][CO1] [12M]**
7	**a)**	Outline the concept of Hadoop configuration file system.	**[L4][CO3] [06M]**
	b)	Describe the Hadoop specification followed by facebook.	**[L1][CO1] [06M]**
8	**a)**	Discuss in detail benchmarking in Hadoop.	**[L2][CO1] [06M]**
	b)	What are the security issues that may arise in the hadoop environment? Explain.	**[L2][CO1] [06M]**
9		Express the steps in installing Hadoop Cluster	**[L2][CO3] [12M]**
10		Analyze the ways of implementing security in Hadoop Environment.	**[L4][CO3] [12M]**

Course Code: 20MC9141 **R20**

UNIT-V

APPLICATIONS ON BIG DATA PIG AND HIVE & VISUALIZATIONS

1	**a)**	What is PIG and explain the process steps.	**[L1][CO5] [06M]**
	b)	Explain two execution types or modes in PIG.	**[L2][CO5] [06M]**
2	**a)**	Describe in detail HIVE services..	**[L1][CO5] [06M]**
	b)	Discuss about the applications on big data using pig & hive.	**[L2][CO5] [06M]**
3	**a)**	Clearly explain the process of installing & features of HIVE.	**[L2][CO5] [06M]**
	b)	Summarize the data processing operators in pig.	**[L2][CO5] [06M]**
4	**a)**	Identify, how will you query the data in HIVE?	**[L1][CO5] [06M]**
	b)	Discuss about Hive Services.	**[L2][CO5] [06M]**
5	**a)**	Briefly discuss about HBASE.	**[L2][CO5] [06M]**
	b)	How to query data in Hive? Explain.	**[L2][CO5] [06M]**
6	**a)**	What is Zookeeper? Explain its features with applications.	**[L2][CO3] [06M]**
	b)	Describe about systems and applications of visualizations	**[L2][CO6] [06M]**
7	**a)**	Explain in detail IBM infosphere Big insights and Streams.	**[L2][CO6] [06M]**
	b)	What are the fundamentals of HBase and Zookeeper?	**[L1][CO5] [06M]**
8	**a)**	Discuss the visual data analysis techniques in detail.	**[L2][CO6] [06M]**
	b)	What are the different types of big data applications?	**[L1][CO5] [06M]**
9	**a)**	Briefly explain Interaction techniques with its applications.	**[L2][CO5] [06M]**
	b)	Match the syntaxes of where, order by, group by & joins	**[L2][CO5] [06M]**
10	**a)**	What is HiveQL? Explain its features.	**[L2][CO5] [06M]**
	b)	Explain the features of HBase in a brief manner.	**[L2][CO5] [06M]**

Prepared by Mr. V. HARSHA VARDHAN, Associate. Professor, Department of MCA, SIETK